

The Medusa of Spatial Sorting: Topological Construction *

Herbert Edelsbrunner[†], Carl-Philipp Heisenberg[‡], Michael Kerber[‡] and Gabriel Krens[‡]

Abstract

We consider the simultaneous movement of finitely many colored points in space, calling it a *spatial sorting process*. The name suggests a purpose that drives the collection to a configuration of increased or decreased order. Mapping such a process to a subset of space-time, we use persistent homology measurements of the time function to characterize the process topologically.

Keywords. Computational topology, persistent homology, alpha complexes, spatial sorting, cell segregation.

1 Introduction

Motivated by the observation of cell communities over time, we propose a topological expression of the process that facilitates the identification and quantification of its features. In this manuscript, we focus on the mathematical framework.

Motivation. The specific motivation for the work described in this paper is the experimental data on cell segregation in the developing zebrafish imaged by Heisenberg and Krens [9]. Making cells of two populations observable through fluorescent markers, they follow them through time, assigning each population (cell type) its own unique color. In this particular case, the two populations start spatially mixed and end in spatially segregated configurations. The segregation is captured by a series of 3-dimensional images, which we turn into a shape in space-time.

*This research is partially supported by NSF under grant DBI-0820624, by ESF under the Research Network Programme, and by the Russian Government under mega project 11.G34.31.0053.

[†]IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria; Departments of Computer Science and of Mathematics, Duke University, Durham, North Carolina; Delaunay Lab of Discrete and Computational Geometry, Yaroslavl' State University, Russia; Geomagic, Research Triangle Park, North Carolina.

[‡]IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria.

Spatial segregation is a special case of the broader class of *spatial sorting processes*, in which we are given one or more distinguishable populations of particles (points in space), and we are interested in characterizing their spatial rearrangement in time. We aim at characterizing the spatial sorting process through detailed measurements of its features. The quantification may be used to establish a classification of spatial sorting processes or, on a finer scale, to differentiate between realizations of the same process. A common biological application is the establishment of phenotypes that can help in the classification of genetic influences. Once we have a description of the process beyond initial and final states, we may ask more subtle questions, such as whether an observed inverse process has a symmetric characterization.

Results. Our contributions are primarily mathematical, with the goal of using the insights toward the quantitative analysis of experimental time-series data:

- (i) we *model* a spatial sorting process as a shape in space-time with descriptive topological properties;
- (ii) we *measure* this shape using the persistent homology of the time function;
- (iii) we *provide* a classification of the measurements, interpreting them as aspects of the process.

Note that measuring the process in space-time is different from taking the trajectory of measurements of the sequence of time-slices. Indeed, we will distinguish between *temporary* space-time features, that can be observed in slices, from *fleeting* features that cannot be so observed. The latter require memory and temporal reasoning and are therefore less readily accessible to an observer who lives in time.

The main idea of our approach is to turn the time-series of geometric data into a 4-dimensional topological space whose connectivity is descriptive of the spatial sorting process. The construction takes three steps to produce the measurements:

STEP 1: construct the Voronoi tessellation to turn the data in a time-slice into a subset of \mathbb{R}^3 ;

STEP 2: maintain the construction through time, effectively sweeping out a subset of $\mathbb{R}^3 \times [0, 1]$;

STEP 3: measure the constructed subset of space-time using the persistent homology of its time function.

We explain the first two steps in Section 2 and the third step in the Section 3. Details of the corresponding algorithms and their implementations can be found in [10]. Section 4 discusses the information provided by persistent homology. Section 5 illustrates the ideas by running the corresponding algorithms on simulated data. Section 6 concludes the paper.

2 Geometry

In this section, we introduce the sets and complexes we use to model spatial sorting processes. In Euclidean space, we call them *tessellations*, *triangulations*, and *complexes*, and in space-time, we call them *medusas*. They all consist of *cells* of various dimensions, but because we also talk about cells in the biological sense, we will use special names, such as *Voronoi cells* and *simplices* in Euclidean space, and *stacks* and *prisms* in space-time, whenever possible.

2.1 A Moment in Time

The input data is a finite set of colored points in \mathbb{R}^3 . Writing the color as subscript, we let U be the union of U_1 to U_k , assuming $U_i \cap U_j = \emptyset$ for all $i \neq j$.

Voronoi tessellation and Delaunay triangulation. The *Voronoi cell* of $u \in U$ consists of all $x \in \mathbb{R}^3$ for which u minimizes the Euclidean distance among all points in U :

$$\text{vor}(u) = \{x \in \mathbb{R}^3 \mid \|x - u\| \leq \|x - v\|, v \in U\}. \quad (1)$$

The set of Voronoi cells, $V = V(U) = \{\text{vor}(u) \mid u \in U\}$, is the *Voronoi tessellation* of U . For each $1 \leq \ell \leq k$, we write $V_\ell = \{\text{vor}(u) \mid u \in U_\ell\}$ for the subset of Voronoi cells of that color. Note that V is the disjoint union of V_1, V_2, \dots, V_k . It is often convenient to work with the dual instead of directly with the Voronoi tessellation. Abstractly, this is the *nerve* of the set of Voronoi cells, that is: the system of subsets whose Voronoi cells have a non-empty common intersection. Assuming general position of the points in U , the number of Voronoi cells that can have a non-empty common intersection is at most 4. We represent each set in the nerve by the convex hull of the points that generate its Voronoi cells, which can be a vertex, and edge, a triangle, or a tetrahedron. Together, these convex hulls form a simplicial complex, known as the *Delaunay triangulation* of U , and denoted

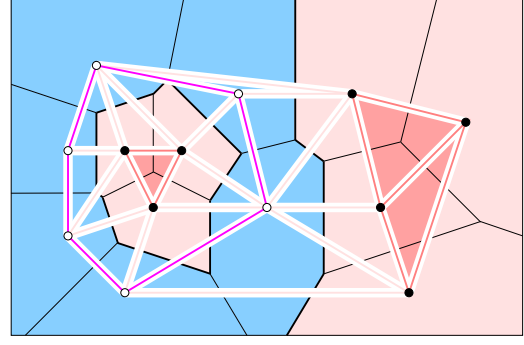


Figure 1: The full subcomplexes defined by the two colors of a bi-chromatic Delaunay triangulation in the Euclidean plane.

as $D = D(U)$. For each $1 \leq \ell \leq k$, we write D_ℓ for the full subcomplex of D that consists of all vertices of color ℓ and all edges, triangles, and tetrahedra that connect them. As illustrated in Figure 1 for \mathbb{R}^2 , the full subcomplexes of different colors are disjoint, with multi-colored edges, triangles, and tetrahedra between them.

Restricted Voronoi tessellation and alpha complex. In a loose configuration, a biological cell would generally not occupy the entire space allotted to it by the Voronoi cell of its nucleus. To better approximate the space used by the cell, we therefore choose a fixed positive radius, α_0 , and consider the *restriction* of the Voronoi cell to the ball centered at the generating point:

$$\text{res}(u) = \{x \in \text{vor}(u) \mid \|x - u\| \leq \alpha_0\}. \quad (2)$$

Similar to before, we write $R = R(U)$ for the set of restricted Voronoi cells, and $R_\ell \subseteq R$ for the subset of cells generated by points of color ℓ . Clearly, R is the disjoint union of R_1, R_2, \dots, R_k . Note that $\text{res}(u) \subseteq \text{vor}(u)$ for each point $u \in U$. It follows that the nerve of R is isomorphic to a subsystem of the nerve of V . Accordingly, we define the dual *alpha complex* to consist of the simplices in D whose corresponding restricted Voronoi cells have a non-empty common intersection, denoting it as $A = A(U)$. For each $1 \leq \ell \leq k$, we again have the full subcomplex $A_\ell \subseteq A$ that contains all vertices of color ℓ and all simplices connecting them. Figure 2 illustrates the definitions by restricting the Voronoi cells in Figure 1 to their disks. Note that A_ℓ is also a subcomplex of D , but not necessarily a full subcomplex. Indeed, we have $A_\ell = D_\ell \cap A$, so A_ℓ is a full subcomplex of D iff $A_\ell = D_\ell$.

Homotopy equivalence. An important structural relationship between unions of Voronoi cells and their dual complexes follows from the Nerve Theorem; see e.g. [6, page

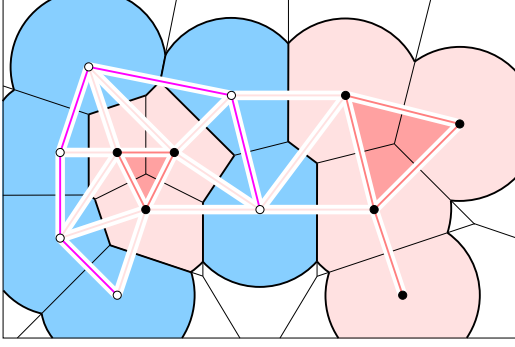


Figure 2: The full subcomplexes defined by the two colors of a bi-chromatic alpha complex in the Euclidean plane.

59]. To state the relationship, we ignore the difference between a set of Voronoi cells and their union, as well as the difference between a simplicial complex and its underlying space. We write $X \simeq Y$ if the sets X and Y have the same homotopy type; see [13, page 108] for a definition.

LEMMA A. We have $R_\ell \simeq A_\ell$, for each color $1 \leq \ell \leq k$.

PROOF. All cells in R_ℓ are closed and convex. It therefore suffices to use the corresponding special version of the Nerve Theorem. It says that the union of the sets and the nerve of the collection have the same homotopy type. The claimed relationships follow because A_ℓ is a geometric realization of the nerves of R_ℓ . \square

We get $V_\ell \simeq D_\ell$ as a special case. Lemma A implies that R_ℓ and A_ℓ have isomorphic homology groups. Instead of dealing with the computationally more difficult union of Voronoi cells, we can therefore compute the homology for the dual complex, which is a purely combinatorial object.

2.2 Trajectories

Assuming a continuous trajectory for each data point, we form subsets of space-time by taking unions of Voronoi cells, both in space and in time. We begin with some definitions. A *trajectory* is a continuous mapping $u : [a, b] \rightarrow \mathbb{R}^3$, with $0 \leq a < b \leq 1$. We let \mathcal{U} be a finite set of trajectories, assuming no two intersect in space-time; that is: $v(t) \neq u(t)$ for all $v \neq u$ in \mathcal{U} and all $t \in [0, 1]$ for which both trajectories are defined. Writing $[k]$ for the set $\{1, 2, \dots, k\}$, we let $\chi : \mathcal{U} \rightarrow [k]$ be a coloring. At each time $t \in [0, 1]$, we have a finite set of points, $\mathcal{U}(t) = \{u(t) \mid u \in \mathcal{U}\}$, of course taking only the trajectories with $a \leq t \leq b$. The finite set of points is also colored, with coloring induced by χ .

Voronoi and Delaunay medusas. For each point $u(t) \in \mathcal{U}(t)$, we write $\text{vor}(u(t))$ for its Voronoi cell in $\mathbb{R}^3 \times t$. The

Voronoi tessellation at time t is denoted as $V(t) = V(\mathcal{U}(t))$, and the subset of Voronoi cells of color ℓ is denoted as $V_\ell(t) \subseteq V(t)$. Collecting Voronoi cells in time, we get a 1-parameter family of cells generated by a trajectory:

$$\text{vor}(u) = \bigcup_{t \in [a, b]} \text{vor}(u(t)). \quad (3)$$

Noting that the Voronoi cells on the right hand side of (3) lie in distinct parallel copies of \mathbb{R}^3 , we call $\text{vor}(u)$ a *stack*. While the Voronoi cell in each time-slice is a 3-dimensional convex polyhedron, the stack itself is neither necessarily convex nor necessarily polyhedral; see Figure 3 on the left. Switching fonts, we write $\mathcal{V} = \mathcal{V}(\mathcal{U})$ for the set of stacks, each defined by a trajectory in \mathcal{U} . For each $1 \leq \ell \leq k$, we write $\mathcal{V}_\ell = \mathcal{V}_\ell(\mathcal{U})$ for the subset of stacks generated by trajectories of color ℓ . We call \mathcal{V} the *multi-chromatic* and each \mathcal{V}_ℓ a *mono-chromatic Voronoi medusa*.

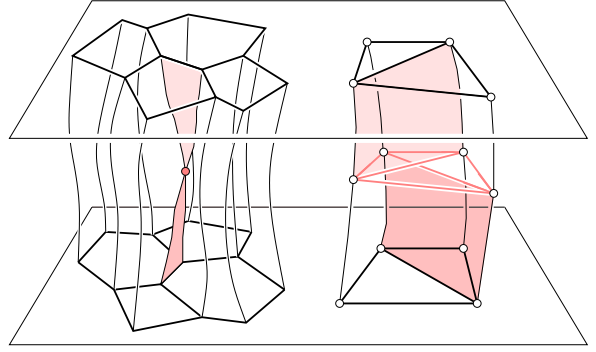


Figure 3: Left: four stack of 2-dimensional Voronoi cells. Right: the four corresponding prisms in the Delaunay medusa connected by a 3-simplex representing the flip between the two ways to triangulate four points.

Similar to stacks of Voronoi cells, we also consider stacks of Delaunay simplices, which we call *prisms*. There is an important difference caused by the occasional sudden change of the Delaunay triangulation. Such a change is a *flip*, which either replaces two tetrahedra by three, or three tetrahedra by two; see e.g. [5, page 102]. In 4-dimensional space-time, a flip appears as a (degenerate) 4-simplex that connects to the preceding Delaunay triangulations along two tetrahedra and to the succeeding Delaunay triangulations along three tetrahedra, or the other way round. Reducing the insertion or deletion of a point to a sequence of flips, as described in [5, Section 5.3], we see that prisms and 4-simplices suffice to fully describe the history of the Delaunay triangulation. We write $\mathcal{D} = \mathcal{D}(\mathcal{U})$ for the complex in $\mathbb{R}^3 \times [0, 1]$ and $\mathcal{D}_\ell = \mathcal{D}_\ell(\mathcal{U})$ for the full subcomplex of color ℓ , calling \mathcal{D} the *multi-chromatic* and \mathcal{D}_ℓ a *mono-chromatic Delaunay medusa*. Figure 3 illustrates the definitions in $\mathbb{R}^2 \times [0, 1]$.

Restricted Voronoi and alpha medusas. Following the distinction between unrestricted and restricted Voronoi cells, we extend the latter to space-time in the obvious way. We thus introduce a *stack* of restricted Voronoi cells,

$$\text{res}(u) = \bigcup_{t \in [a, b]} \text{res}(u(t)), \quad (4)$$

the set of such stacks, $\mathcal{R} = \mathcal{R}(\mathcal{U})$, and the complex of prisms and 4-simplices swept out by the simplices in the alpha complex, $\mathcal{A} = \mathcal{A}(\mathcal{U})$. Furthermore, we introduce the colored subsets, $\mathcal{R}_\ell \subseteq \mathcal{R}$, and the colored subcomplexes, $\mathcal{A}_\ell \subseteq \mathcal{A}$, with $\ell \in [k]$. We call \mathcal{R} the *multi-chromatic* and each \mathcal{R}_ℓ a *mono-chromatic restricted Voronoi medusa*. Similarly, we call \mathcal{A} the *multi-chromatic* and each \mathcal{A}_ℓ a *mono-chromatic alpha medusa*.

Homotopy equivalence. The structural relationship between unions of Voronoi cells and their dual complexes extends from \mathbb{R}^3 to $\mathbb{R}^3 \times [0, 1]$. As before, we simplify the notation by ignoring the difference between a collection of cells and its union.

LEMMA B. We have $\mathcal{R}_\ell \simeq \mathcal{A}_\ell$, for each color $\ell \in [k]$.

PROOF. As in the proof of Lemma A, we appeal to the Nerve Theorem to give the claimed relationships. Since the cells are no longer convex, we use the more general version that applies to finite collections of closed, contractible sets whose intersections (of any order) are again contractible.

While each stack in \mathcal{R}_ℓ of any dimension is swept out by a convex set and is therefore contractible, two stacks can intersect in two or more (lower-dimensional) stacks, which prevents the direct application of the Nerve Theorem. We finesse the difficulty by shrinking each stack at its boundary and thickening it into a 4-dimensional body; see Figure 4. There is more than one way to do this such that the bodies are closed and contractible, their union is the same as the union of the original stacks, and the common intersection of $i + 1$ bodies is either empty or a contractible $(4 - i)$ -dimensional set. For example, we may use the mixed complex defined for Voronoi cells in [4] to sweep out the bodies. Letting \mathcal{B}_ℓ be the resulting set of bodies, we consider the nerve, which we denote by $\bar{\mathcal{B}}_\ell$. As illustrated in Figure 4, this nerve is the barycentric subdivision of a complex of simplices whose vertices correspond to the original 4-dimensional stacks. Note that this complex is *not* necessarily a simplicial complex, since two simplices may share more than just one common face. Denoting this complex of simplices by $\bar{\mathcal{A}}_\ell$, we have

$$\mathcal{R}_\ell \simeq \mathcal{B}_\ell \simeq \bar{\mathcal{B}}_\ell \simeq \bar{\mathcal{A}}_\ell, \quad (5)$$

in which the only non-trivial, middle relation is furnished by the Nerve Theorem. To complete the argument, we still need

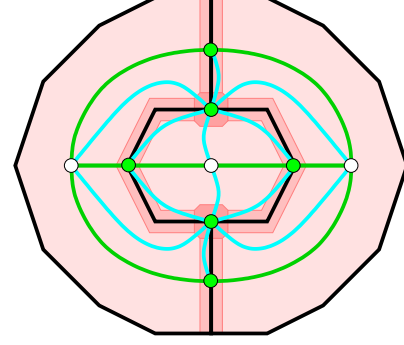


Figure 4: Three 2-dimensional stacks (lightly shaded) giving rise to a decomposition into nine bodies (shaded). The nerve of the bodies forms the barycentric subdivision of two triangles.

to show that $\bar{\mathcal{A}}_\ell$ and \mathcal{A}_ℓ have the same homotopy type. To do this, we contract each prism in \mathcal{A}_ℓ along the time direction. Indeed, each prism is a simplex times a time interval, and the contraction glues the bottom to the top face, effectively turning the prism into a simplex. We do this for all prisms simultaneously in order to avoid temporary inconsistencies between the prisms and their side faces, which are lower-dimensional prisms. Eventually, when all prisms have been contracted, we have turned the medusa into $\bar{\mathcal{A}}_\ell$. Since the contraction preserves the homotopy type, this implies the claimed relation. \square

We get $\mathcal{V}_\ell \simeq \mathcal{D}_\ell$ as a special case. Lemma B also implies that we get isomorphic homology groups and can therefore do the computation on the dual medusa, or indeed the complex of simplices, $\bar{\mathcal{A}}_\ell$, which is a purely combinatorial object. We refer to it as the *data structure* for the mono-chromatic alpha medusa; beyond being instrumental in the proof of Lemma B, it is the main computational tool we use to implement that algorithms that are only sketched in this paper; see [10].

2.3 Filtered Sequences

We use the restricted Voronoi medusa to explain filtered sequences of spaces, and free up to subscript by focusing the discussion on the multi-chromatic version.

Sub- and superlevel sets. The appropriately restricted *time function*, $f|_{\mathcal{R}} : \mathcal{R} \rightarrow [0, 1]$, maps every point $x \in \mathcal{R}$ to its time coordinate, $f|_{\mathcal{R}}(x)$. Given a moment in time, $t \in [0, 1]$, the corresponding *sublevel set* is $\mathcal{R}_t = f|_{\mathcal{R}}^{-1}[0, t]$, and the corresponding *superlevel set* is $\mathcal{R}^t = f|_{\mathcal{R}}^{-1}[t, 1]$. We will be interested in the filtered sequences:

$$\emptyset \subseteq \mathcal{R}_0 \subseteq \dots \subseteq \mathcal{R}_t \subseteq \dots \subseteq \mathcal{R}_1 = \mathcal{R}, \quad (6)$$

$$\emptyset \subseteq \mathcal{R}^1 \subseteq \dots \subseteq \mathcal{R}^t \subseteq \dots \subseteq \mathcal{R}^0 = \mathcal{R}. \quad (7)$$

They capture the evolution of the restricted Voronoi medusa in a way that allows us to compare different states without the burden of computing a correspondence. Similarly, we consider the time function restricted to the alpha medusa, define sub- and superlevel sets, \mathcal{A}_t and \mathcal{A}^t , and form filtered sequences as before. Importantly, the transformation from the stacks to the prisms does not change the homotopy type.

LEMMA C. We have $\mathcal{R}_t \simeq \mathcal{A}_t$ as well as $\mathcal{R}^t \simeq \mathcal{A}^t$, for all $t \in [0, 1]$.

We omit the proof which is similar to that of Lemma B. It is important to realize that Lemma C extends to all monochromatic medusas.

Complexes of simplices. Following the proof of Lemma B, we find it convenient to replace the sub- and superlevel sets of the alpha medusa by complexes of simplices. Specifically, we write $\bar{\mathcal{A}}_t$ for the complex obtained by contracting all prisms in \mathcal{A}_t in time direction. Symmetrically, we write $\bar{\mathcal{A}}^t$ for the complex obtained by contracting all prisms in \mathcal{A}^t . There is an alternative description. Let φ be a prism in \mathcal{A} , and write $f_{\min}(\varphi)$ and $f_{\max}(\varphi)$ for the minimum and maximum time values of the points in φ . We assign these two values as $\sigma_{\min} \leq \sigma_{\max}$ to the simplex σ in $\bar{\mathcal{A}}$ obtained by contracting φ . Then $\bar{\mathcal{A}}_t$ is the subcomplex of simplices with $\sigma_{\min} \leq t$, and $\bar{\mathcal{A}}^t$ is the subcomplex of simplices with $\sigma_{\max} \geq t$. Note that $\bar{\mathcal{A}}_t$ and $\bar{\mathcal{A}}^t$ overlap in the simplices that correspond to the prisms with $f_{\min}(\varphi) \leq t \leq f_{\max}(\varphi)$. Similar to the sub- and superlevel sets, the complexes of simplices form filtered sequences:

$$\emptyset \subseteq \bar{\mathcal{A}}_0 \subseteq \dots \subseteq \bar{\mathcal{A}}_t \subseteq \dots \subseteq \bar{\mathcal{A}}_1 = \bar{\mathcal{A}}, \quad (8)$$

$$\emptyset \subseteq \bar{\mathcal{A}}^1 \subseteq \dots \subseteq \bar{\mathcal{A}}^t \subseteq \dots \subseteq \bar{\mathcal{A}}^0 = \bar{\mathcal{A}}, \quad (9)$$

Again, the transformation does not affect the homotopy type. Indeed, the complexes forming the filtered sequences in (8) and (9) are reminiscent of the lower- and upper-star filtrations we find in [6].

3 Algebra

In this section, we turn the spaces of Section 2 into algebraic information. The foundation of the transformation is the classical notion of homology, which we review. The information is summarized in persistence diagrams, which we introduce for modules obtained from sub- and superlevel sets of the time function.

3.1 Measuring Connectivity

Homology groups detect and count holes in a single space. We begin with a brief introduction of this classical subject; see [13] for more information.

Homology groups. Consider a simplicial complex, perhaps an alpha complex, A , which consists of simplices of dimension $0 \leq p \leq 3$. Fixing a field of coefficients, F , we call a formal sum of the form $\gamma = \sum c_i \sigma_i$ a p -chain, in which the c_i are elements of F and the σ_i are p -simplices in A , each with a fixed orientation. The *boundary* of the p -chain is the similarly weighted sum of boundaries of the simplices: $\partial_p \gamma = \sum c_i \partial_p \sigma_i$, in which $\partial_p \sigma_i$ is the sum of the $(p-1)$ -simplices that are its faces. We call γ a p -cycle if $\partial_p \gamma = 0$, and we call γ a p -boundary if there is a $(p+1)$ -chain δ with $\gamma = \partial_{p+1} \delta$. The chains thus form *chain groups* connected by boundary homomorphisms, $\partial_p : C_p \rightarrow C_{p-1}$. Similarly, the cycles form *cycle groups*, the kernels of the boundary homomorphisms, $Z_p = \ker \partial_p$, and the boundaries form *boundary groups*, the images of the boundary homomorphisms, $B_p = \text{im } \partial_{p+1}$. Since the boundary of a boundary is necessarily zero, we have $B_p \subseteq Z_p$ and we can take the quotient, $H_p = Z_p / B_p$, which is the p -th *homology group*. Homology groups can be defined quite generally, for example, as explained above for triangulations of topological spaces. Since we choose the coefficients from a field, all groups we mentioned are vector spaces, which are characterized by their ranks. For the p -th homology group, the rank is called the p -th *Betti number*, denoted as $\beta_p = \text{rank } H_p$.

For a space of dimension k , the only possibly non-zero Betti numbers are β_0 to β_k . In our case, A has dimension at most $k = 3$, and we have $\beta_3 = 0$ because every 3-chain in \mathbb{R}^3 has non-zero boundary. The remaining three possibly non-zero Betti numbers have intuitive interpretations: β_0 counts components, β_1 counts loops, and β_2 counts completely surrounding walls. We get additional intuition by observing that the connectivity of the complement space, $\mathbb{R}^3 - A$, depends on the connectivity of A , a relation formalized by Alexander Duality [13, p. 424]. We refer to the elements of the homology group of the complement space as the *holes* of A . Distinguishing between the different dimensions, $\beta_0 - 1$ counts *gaps* between the components, β_1 counts *tunnels* passing through the loops, and β_2 counts *voids* surrounded by walls. We will compute Betti numbers for medusas in $\mathbb{R}^3 \times [0, 1]$. While more complicated than in 3-dimensional space, we can still interpret the Betti numbers in terms of *evolutions* of gaps, tunnels, and voids, as we will explain in Section 4.

In addition to single spaces, we take the homology of pairs in which the second space is a subset of the first, such as $(\mathcal{R}, \mathcal{R}^t)$, for example. The p -th *relative homology group*, denoted as $H_p(\mathcal{R}, \mathcal{R}^t)$, is defined the same way as $H_p(\mathcal{R})$ except that differences in \mathcal{R}^t are ignored. In other words, two chains are the same if they agree on all simplices in $\mathcal{R} - \mathcal{R}^t$, etc. To get an intuition for these groups, we may add the cone over \mathcal{R}^t . Then the rank of $H_p(\mathcal{R}, \mathcal{R}^t)$ is the same as the rank of the p -th absolute homology group for the modified complex, except for $p = 0$, where it is one less.

Ordinary and extended persistence modules. In the language of category theory, homology is a functor that maps a space, or a pair of spaces, to a sequence of groups, one for each dimension. To simplify the notation, we write $H = \bigoplus_p H_p$ for the direct sum of the individual groups. This frees up the subscript, which we use to index the homology groups of a filtered sequence. Letting $0 = t_1 < t_2 < \dots < t_m = 1$ be the homological critical values of $f|_{\mathcal{R}}$, we note that the homology group is constant for all $t_i \leq t < t_{i+1}$. Hence, it suffices to list the finitely many groups $R_i = H(\mathcal{R}_{t_i})$ to describe the entire 1-parameter family, giving a finite sequence of homology groups connected by homomorphisms induced by inclusion:

$$0 = R_0 \rightarrow R_1 \rightarrow \dots \rightarrow R_m = H(\mathcal{R}). \quad (10)$$

We call (10) a *persistence module*, or more specifically the *ordinary persistence module* of f . As described in [6, Section VII.1], we can use the persistence module to define *births* and *deaths* of homology classes and arrange them in pairs. Specifically, a class is *born* at R_i if it does not belong to the image of R_{i-1} , and it *dies entering* R_j if j is the smallest index for which the image of the class lies in the image of R_{i-1} . The birth-death pair describes a coset of homology classes and specifies their *persistence* as the absolute difference between the function values at the two events, in this case $t_j - t_i$.

The ordinary module (10) allows for homology classes that are born but never die. These are the classes that describe the entire medusa, \mathcal{R} , which are of particular interest to us. To get a finite measurement of duration, we extend the module using the relative homology groups of the form $H(\mathcal{R}, \mathcal{R}^t)$. Writing $R_{m+i} = H(\mathcal{R}, \mathcal{R}^{t_{m-i+1}})$, we get

$$0 = R_0 \rightarrow \dots \rightarrow R_m \rightarrow \dots \rightarrow R_{2m} = 0, \quad (11)$$

which we call the *extended persistence module* of $f|_{\mathcal{R}}$. It begins and ends with the trivial group, which implies that everything that is born also dies. We thus get a more complete set of measurements of the alpha medusa from (11) than from (10).

Isomorphic relative homology groups. Lemma C implies that the (absolute) homology groups of the sublevel sets of the restricted Voronoi medusa are isomorphic to those of the alpha medusa. We now show that the same is true for the relative homology groups.

LEMMA D. The relative homology groups $H(\mathcal{R}, \mathcal{R}^t)$ and $H(\mathcal{A}, \mathcal{A}^t)$ are isomorphic, for all $t \in [0, 1]$.

PROOF. Recall the exact sequences of the two pairs, which we write from left to right and in parallel:

$$\begin{array}{ccccccc} H_p(\mathcal{A}^t) & \rightarrow & H_p(\mathcal{A}) & \rightarrow & H_p(\mathcal{A}, \mathcal{A}^t) & \rightarrow & H_{p-1}(\mathcal{A}^t) \rightarrow H_{p-1}(\mathcal{A}) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ H_p(\mathcal{R}^t) & \rightarrow & H_p(\mathcal{R}) & \rightarrow & H_p(\mathcal{R}, \mathcal{R}^t) & \rightarrow & H_{p-1}(\mathcal{R}^t) \rightarrow H_{p-1}(\mathcal{R}). \end{array}$$

By Lemma C, the vertical maps between the first, second, fourth, and fifth groups are isomorphisms. The diagram commutes because all maps are induced by inclusion. The Steenrod Five Lemma thus implies that the vertical maps between the middle groups is also an isomorphism [13, p. 140]. \square

Similar to Lemma C, Lemma D extends to the monochromatic case.

3.2 Persistent Homology

It is instructive to display the information contained in a persistence module as a finite multi-set of points (referred to as *dots*) in the plane. After explaining how this is done, we prove that the sequences of homotopy equivalent spaces introduced above give the same diagram.

Persistence diagram. As before, we consider the restricted Voronoi medusa, \mathcal{R} , the time function $f|_{\mathcal{R}} : \mathcal{R} \rightarrow [0, 1]$, and the extended module defined by the sub- and superlevel sets. The *persistence diagram* of $f|_{\mathcal{R}}$, which we denote as $\text{Dgm}(f|_{\mathcal{R}})$, is a multi-set of dots in a double covering of the plane; see Figure 5. Each dot has two coordinates and represents a coset of homology classes. These classes have a homological dimension, which we use to label the dot. Very often, we consider subdiagrams by limiting ourselves to dots of a particular dimension. For example, $\text{Dgm}_0(f|_{\mathcal{R}})$ is the multiset of dots of homological dimension 0, which therefore only represents components. The arrows in our drawing indicate the direction of increasing time. Since we have two phases, one sweeping forward and the other backward in time, we get every pair of coordinates twice. To explain this in more detail, we call the two coordinates the *birth* and the *death* of the dot (or of the homology classes it represents). The birth axes appear as -45° lines in the figure, and the death axes as 45° lines. We decompose the persistence

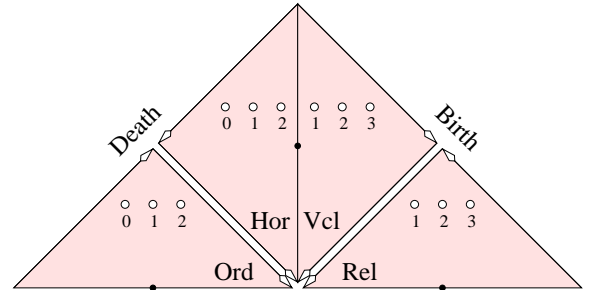


Figure 5: From left to right: the ordinary, the horizontal, the vertical, and the relative subdiagram of the persistence diagram. The types of dots that can exist in the diagram of a medusa in 4-dimensional space-time are shown.

diagram into four triangles, referred to as the *ordinary*, *horizontal*, *vertical*, *relative subdiagrams*, denoted as $\text{Ord}(f|_{\mathcal{R}})$, $\text{Hor}(f|_{\mathcal{R}})$, $\text{Vcl}(f|_{\mathcal{R}})$, $\text{Rel}(f|_{\mathcal{R}})$. A dot in $\text{Ord}(f|_{\mathcal{R}})$ represents classes that are born and die during the first phase, a dot in $\text{Hor}(f|_{\mathcal{R}})$ or $\text{Vcl}(f|_{\mathcal{R}})$ represents classes that are born during the first phase and die during the second phase, and a dot in $\text{Rel}(f|_{\mathcal{R}})$ represents classes that are born and die during the second phase. The difference between a dot in $\text{Hor}(f|_{\mathcal{R}})$ and in $\text{Vcl}(f|_{\mathcal{R}})$ is that for the former the birth coordinate is smaller than the death coordinate, while for the latter it is the other way round. We believe this difference is important, as we will explain later.

Equivalence. Using Lemmas A to D, we can now make sure that the diagrams we get from the different medusas are the same. As before, we limit the discussion to the multi-chromatic restricted Voronoi and alpha medusas. Let $f|_{\mathcal{R}}$ and $f|_{\mathcal{V}}$ be the corresponding time functions.

LEMMA E. $\text{Dgm}(f|_{\mathcal{A}})$ and $\text{Dgm}(f|_{\mathcal{R}})$ are the same.

PROOF. The main tool in this proof is the diagram that connects the homology groups in the two modules:

$$\begin{array}{ccccccc} H(\mathcal{A}_0) & \rightarrow & \dots & \rightarrow & H(\mathcal{A}_1) & \rightarrow & H(\mathcal{A}, \mathcal{A}^1) \rightarrow \dots \rightarrow H(\mathcal{A}, \mathcal{A}^0) \\ \downarrow & & & & \downarrow & & \downarrow \\ H(\mathcal{R}_0) & \rightarrow & \dots & \rightarrow & H(\mathcal{R}_1) & \rightarrow & H(\mathcal{R}, \mathcal{R}^1) \rightarrow \dots \rightarrow H(\mathcal{R}, \mathcal{R}^0). \end{array}$$

By Lemmas C and D, the vertical maps are isomorphisms. Since the diagram commutes, the claim follows by the Persistence Equivalence Theorem in [6, p. 159]. \square

Similar to Lemmas C and D, Lemma E extends to the mono-chromatic case, and it implies the same relation for Delaunay and Voronoi medusas as a special case.

3.3 Images

We expect the measurements for the alpha medusa to be more meaningful than for the Delaunay medusa, but there is additional information in the difference. We compare by studying the images of the homomorphisms induced by the inclusion of the one in the other medusa. While we use this ability for mono-chromatic medusas, we simplify the notation by describing it in the multi-chromatic case.

Image persistence. Recall that \mathcal{R}_t , \mathcal{R}^t , \mathcal{V}_t , and \mathcal{V}^t are the sublevel and superlevel sets of the Voronoi medusas, for $t \in [0, 1]$. By construction, we have $\mathcal{R}_t \subseteq \mathcal{V}_t$ and $\mathcal{R}^t \subseteq \mathcal{V}^t$. Applying the homology functor, we get finitely many distinct groups, which we denote as R_i for the restricted and as V_i for the unrestricted Voronoi medusa. Aligning the two

modules, we repeat groups as necessary and arrange them in a commuting diagram:

$$\begin{array}{ccccccc} R_0 & \rightarrow & R_1 & \rightarrow & \dots & \rightarrow & R_N \\ \downarrow & & \downarrow & & & & \downarrow \\ V_0 & \rightarrow & V_1 & \rightarrow & \dots & \rightarrow & V_N. \end{array}$$

Writing $\iota_i : R_i \rightarrow V_i$ for the homomorphism induced by the inclusion, we are interested in the sequence of images:

$$0 = \text{im } \iota_0 \rightarrow \text{im } \iota_1 \rightarrow \dots \rightarrow \text{im } \iota_N = 0. \quad (12)$$

Similar to the module of homology groups, (12) is a sequence of vector spaces connected by homomorphisms. We can therefore define births and deaths. We refer to the corresponding multiset of dots as the *image persistence diagram*, denoted as $\text{Dgm}(\text{im } f|_{\mathcal{R}} \rightarrow f|_{\mathcal{V}})$; see [2] for a detailed discussion of this construction and for an algorithm.

What does the diagram measure? In the assumed case of the multi-chromatic restricted included in the multi-chromatic unrestricted Voronoi medusa, it measures nothing interesting, simply because the groups V_i are not interesting. This is different in the mono-chromatic case. Here, we may have a cycle defined by data points of color ℓ surrounding points of color different from ℓ . In this case, we have a non-trivial class in a (closed or open) sublevel set of the restricted Voronoi medusa whose image in the corresponding sublevel set of the unrestricted Voronoi medusa is still non-trivial. Indeed, a cycle in \mathcal{R}_ℓ gives rise to a dot in $\text{Dgm}(\text{im } f|_{\mathcal{R}_\ell} \rightarrow f|_{\mathcal{V}_\ell})$ iff it corresponds to a hole formed by points of color different from ℓ . If the hole is not formed by such points, then it does not exist in the unrestricted Voronoi medusa, the class maps to 0, and there is no corresponding dot in the image persistence diagram.

Instead of $\mathcal{R}_\ell \subseteq \mathcal{V}_\ell$, we can use the inclusion of \mathcal{R}_ℓ in \mathcal{R} to recognize when holes are caused by interactions between different colors. The algebraic set-up is the same, so we do not need to repeat it. As described in the experimental Section 5, the latter inclusion seems to be more effective than the inclusion in the mono-chromatic Voronoi medusa. This is perhaps related to the fact that the most interesting is also the most difficult case in this respect, namely that of a 1-dimensional homology class in \mathbb{R}^3 .

Equivalence. Similar to the conventional case, the image persistence diagrams do not depend on the representation of the space and the time function we use. Specifically, we get the same diagrams for the inclusion of \mathcal{R} in \mathcal{V} as for the inclusion of \mathcal{A} in \mathcal{D} .

LEMMA F. $\text{Dgm}(\text{im } f|_{\mathcal{R}} \rightarrow f|_{\mathcal{V}})$ and $\text{Dgm}(\text{im } f|_{\mathcal{A}} \rightarrow f|_{\mathcal{D}})$ are equal as diagrams.

PROOF. Arrange the four modules in a 3-dimensional diagram, in which the 2-dimensional section at time t consists of the (absolute) homology groups of the sublevel sets:

$$\begin{array}{ccc} H(\mathcal{A}_t) & \rightarrow & H(\mathcal{R}_t) \\ \downarrow & & \downarrow \\ H(\mathcal{D}_t) & \rightarrow & H(\mathcal{V}_t), \end{array}$$

or of the relative homology groups if the section is taken during the second phase. In all three directions, the maps are induced by inclusion, so the diagram commutes. This implies that we have maps between the corresponding two modules of images. By Lemmas C and D, the horizontal maps in the 3-dimensional diagram are isomorphisms, which implies that the maps connecting the two modules of images are isomorphisms. The claimed relation is implied by the Persistence Equivalence Theorem. \square

Similar to Lemmas C, D, and E, Lemma F extends to the mono-chromatic case. We note that the two derived persistence diagrams are best computed using the complexes of simplices, $\bar{\mathcal{A}}$ and $\bar{\mathcal{D}}$. Indeed, these can be connected by a mapping cylinder, giving a complex of simplices and prisms, not unlike but different from the alpha medusa. From this complex, we get the image persistence diagrams by standard matrix reduction; see [2].

4 Classification

In this section, we discuss the information contained in the extended persistence module defined by the time function, interpreting the corresponding events in static space-time as well as in dynamic temporal language.

4.1 Plane \times Time

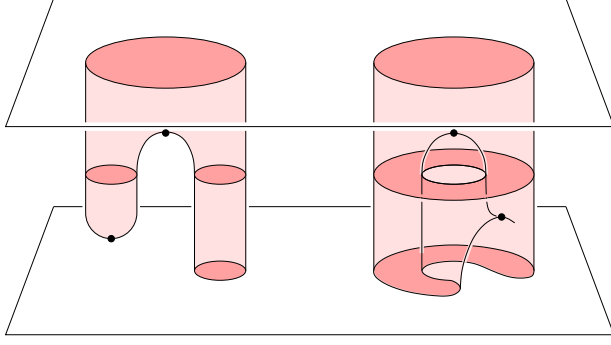
We begin with the 2-dimensional case as a warm-up exercise, but also to facilitate the comparison with the 3-dimensional case. Here, the medusa is embedded in $\mathbb{R}^2 \times [0, 1]$, and the time function, denoted as f , maps each point to its time coordinate. We can have dots in the diagram for dimensions 0, 1, 2, except for $\text{Ord}_2(f)$, $\text{Hor}_2(f)$, $\text{Vcl}_0(f)$, and $\text{Rel}_0(f)$. Recall that the persistence of the class represented by a dot $X = (a, b)$ is the absolute difference between the two coordinates, which we denote as $\text{pers}(X) = |a - b|$. In the ordinary and relative subdiagrams, $\text{pers}(X)$ is $\sqrt{2}$ times the distance from the baseline, while in the horizontal and vertical subdiagrams, it is $\sqrt{2}$ times the distance from the vertical axis that goes through the middle of the diagram. As established in Section 3, we have the choice between several filtrations, all giving the same persistence diagram. To focus

the discussion, we limit ourselves to the medusa that best approximates the reality of biological cells, namely the mono-chromatic restricted Voronoi medusa. In the text below, we use *cell* in the biological meaning of the word.

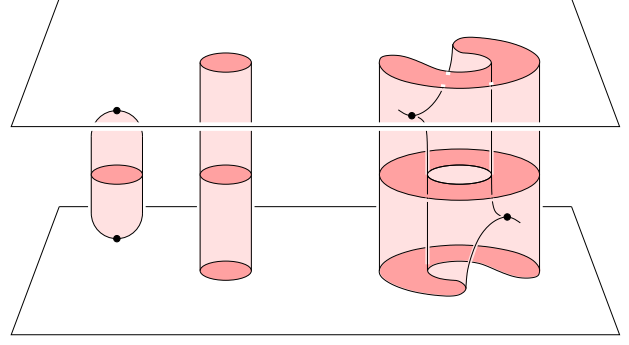
Type analysis. The classes recorded in the ordinary subdiagram are illustrated in Figure 6 (a). A dot $X = (a, b)$ in $\text{Ord}_0(f)$ corresponds to a component born at $t = a$ that merges with another, older component at $t = b$. If $a = 0$, then a non-empty subset of the cells that make up the component exist at the very beginning of the observed time-series. Its death marks the merger with another component also born at 0, and the tie in the decision who dies and who lives is broken arbitrarily. If $a > 0$, the component is born because of a new cell that is either created by cell division or enters the observation for technical reasons. A dot $X = (a, b)$ in $\text{Ord}_1(f)$ corresponds to a ring (an annulus) that forms at $t = a$ and whose hole fills up at $t = b$.

The classes in the horizontal subdiagram are illustrated in Figure 6 (b). A dot $X = (a, b)$ in $\text{Hor}_0(f)$ corresponds to a component born on the way up at $t = a$ and dying on the way down at $t = b$. We have $a < b$ since a is the minimum and b is the maximum time value of the points in the component. Assuming some cells of color ℓ live from the beginning to the end of the observations, we have at least one dot with coordinates 0 and 1 in $\text{Hor}_0(f)$. In the case of a successful segregation, in which color ℓ eventually forms a single component, we have exactly one such dot and no other other dots with second coordinate equal to 1 in the 0-dimensional horizontal subdiagram. A dot $X = (a, b)$ in $\text{Hor}_1(f)$ corresponds to a ring that forms at $t = a$ and breaks up at $t = b$. The breaking up of the ring is not detected by the homology of the growing sublevel set, but rather later by the relative homology during the down phase. The classes recorded in the vertical subdiagram are illustrated in Figure 6 (c). A dot $X = (b, a)$ in $\text{Vcl}_1(f)$ corresponds to a tunnel born on the way up at $t = b$ which dies on the way down at $t = a$, where $a < b$. The tunnel is the plane-time expression of a gap that temporarily opens up between two subpopulations of a component that later closes again. A dot $X = (b, a)$ in $\text{Vcl}_2(f)$ corresponds to a void born on the way up at $t = b$ which dies on the way down at $t = a$, where $a < b$. The void is formed by a temporary ring structure we see within an interval of level sets. In contrast to the rings discussed earlier, this ring is formed by puncturing. First, the hole expands until it eventually shrinks back to a point and disappears at $t = b$.

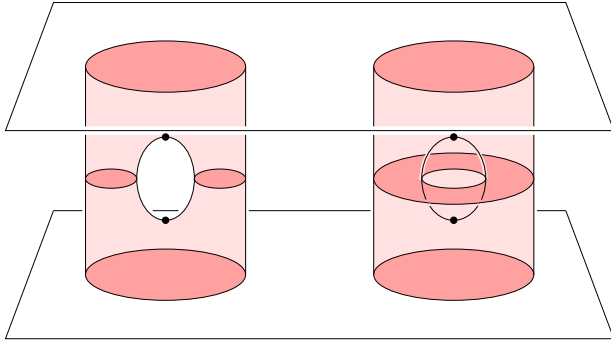
The classes recorded in the relative subdiagram are illustrated in Figure 6 (d). A dot $X = (b, a)$ in $\text{Rel}_1(f)$ corresponds to a component of the level set that splits off another component at $t = a$, and dies off at $t = b$. Since this event is detected in the down phase of the module, we see the birth of a 1-dimensional class at $t = b$ and its death at $t = a$. A



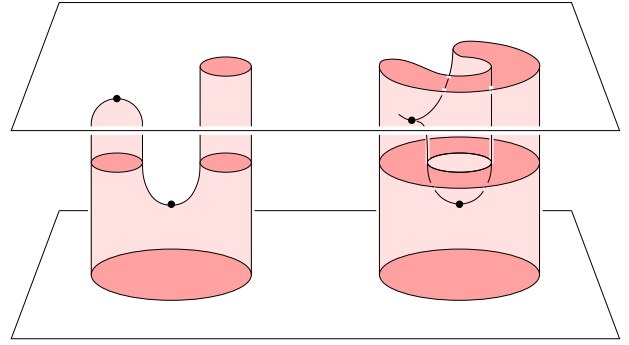
(a) Ordinary subdiagram. Left: a gap appears when a new component is born and disappears when this component merges with another component. Right: a ring appears when two ends of a component meet and disappears when the surrounded hole fills up.



(b) Horizontal subdiagram. Left: a gap appears and disappears when a new component is formed and ends. Right: a ring appears when two ends of a component meet and disappears when the ring breaks open.



(c) Vertical subdiagram. Left: a gap appears when a component splits in two and disappears when the two merge back together into a single component. Right: a ring appears when a disk is punctured, which creates a hole that first grows and later fills up.



(d) Relative subdiagram. Left: a gap appears when a new component is split off and disappears when the component ends. Right: a ring appears when the disk is punctured and disappears when the ring breaks open.

Figure 6: The classes recorded in the four subdiagrams of the persistence diagram.

dot $X = (b, a)$ in $\text{Rel}_2(f)$ corresponds to ring formed by puncturing a disk at $t = a$. The hole expands and eventually disappeared because of the breaking up of the ring at $t = b$.

Interaction between colors. The appearance and disappearance of a hole may or may not be facilitated by interactions with cells of different color. For example, a gap recorded in the ordinary subdiagram may disappear by squeezing out cells of color different from ℓ , or simply by locally consolidating the configuration. In the former case, it is likely that the gap also existed in \mathcal{V}_ℓ and unlikely that it existed in \mathcal{R} . Accordingly, the gap in \mathcal{R}_ℓ would be represented by a dot in the image persistence diagram of $\mathcal{R}_\ell \subseteq \mathcal{V}_\ell$ but not in that of $\mathcal{R}_\ell \subseteq \mathcal{R}$. In the latter case, it would of course be the other way round. Similar distinctions can be made for gaps recorded in Hor_0 , Vcl_1 , and Rel_1 . The analysis of rings formed by cells of color ℓ is similar. A ring may form around a population of cells of color different from ℓ , or around empty space. In the former case, the ring can only disappear by killing the surrounded cells or by breaking up

to release these cells. Since killing seems unlikely, this may mean that such rings are rare in the ordinary and more common in the horizontal subdiagram. Similarly, the latter case is likely to be the preferred configuration for rings recorded in the vertical and the relative subdiagrams. Indeed, these rings form by puncturing and can therefore not be invaded by cells from the outside. Note, however, that the suggestion that a dot in the persistence diagram of \mathcal{R}_ℓ either appears or does not appear in the image persistence diagram is an oversimplification. Thinking of a dot as an interval, it can break up into shorter intervals, or merge with others into a longer interval. Similarly, there is generally no simple relationship between the image persistence diagram and the diagram of the containing medusa.

4.2 Space \times Time

Similar to the planar case, we discuss the time function on a mono-chromatic restricted Voronoi medusa, $f : \mathcal{R}_\ell \rightarrow [0, 1]$ with $\mathcal{R}_\ell \subseteq \mathbb{R}^3 \times [0, 1]$.

Types. Compared to the planar case, we have one more dimension and therefore one additional case in each subdiagram; see Figure 5. In temporal language, we follow the evolution of 0-, 1-, and 2-dimensional holes, which we refer to as *gaps*, *tunnels*, and *voids*. There are two ways a hole can appear: by *closing* the surrounding cycle, or by *puncturing*. Of course, a hole can already exist at the beginning, at $t = 0.0$, in which case its birth coordinate is 0.0. Similarly, there are two ways a hole can disappear: by *breaking up* the surrounding cycle, or by *filling up* the hole. In addition, it may remain to the end, at $t = 1.0$, in which case the death coordinate is 1.0. We observe that a p -dimensional hole in the level set gives rise to either a p -dimensional or a $(p + 1)$ -dimensional class in the persistence module, and which it is depends on how the hole comes into existence: either by closing a p -dimensional cycle or by puncturing a $(p + 1)$ -dimensional chain; see Table 1. A hole recorded in the ordinary subdiagram is formed by closing a cycle, allowing for the case that it exists already at the beginning, and it disappears by filling up. We refer to this action as *aggregation*. Symmetrically, a hole recorded in the relative subdiagram is formed by puncturing, and the surrounding cycle either breaks up or the hole remains until the end. We refer to this action as *disaggregation*. In contrast, the holes recorded in the vertical and horizontal subdiagrams represent aggregations and disaggregations that are either transient or incomplete. Specifically, a hole recorded in the vertical subdiagram appears by puncturing, but instead of breaking up, it disappears by filling. Finally, a hole in the horizontal subdiagram forms by closing a cycle that breaks up later, which allows for the case that the hole exists already at the beginning or that it remains to the end.

	Ord	Hor	Vcl	Rel
	close/fill	close/break	punct/fill	punct/break
0	gap	gap		
1	tunnel	tunnel	gap	gap
2	void	void	tunnel	tunnel
3			void	void

Table 1: The twelve cases that arise from four kinds of evolutions possible for each of three different types of holes in a 1-parameter family of spaces in \mathbb{R}^3 .

Rings and tunnels. The 0- and 2-dimensional holes in \mathbb{R}^3 behave like the 0- and 1-dimensional holes in \mathbb{R}^2 , while the tunnels introduce behavior that cannot be observed in the planar case. Drawing the ring passing around a tunnel as a solid torus, Figure 7 shows the four different evolutions recorded by dots in the four subdiagrams of $\text{Dgm}(f)$.

As in the other cases, it is interesting to distinguish between tunnels formed by interactions with cells of color dif-

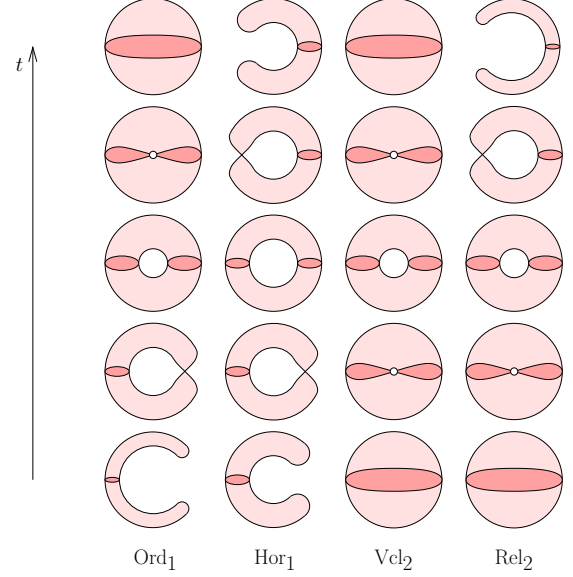


Figure 7: From left to right: the evolution of a ring/tunnel with recording dot in the ordinary, horizontal, vertical, relative subdiagram of $\text{Dgm}(f)$. The critical configurations in the second and fourth rows separate the regular configurations in the other rows.

ferent from ℓ , from tunnels formed without such interactions. As before, we use the image persistence diagrams induced by $\mathcal{R}_\ell \subseteq \mathcal{V}_\ell$ or $\mathcal{R}_\ell \subseteq \mathcal{R}$. If the tunnel corresponds to a dot in the diagram induced by the inclusion in \mathcal{V}_ℓ , then this is evidence for a different color cell squeezing through an opening in the wall of color ℓ . Similar evidence for this event is provided by the absence of this dot in the diagram induced by the inclusion in \mathcal{R} . The different color cell may be successful and move from one side of the wall to the other, or it may be repelled by the force holding the wall together and bounce back. Which case it is may sometimes be apparent but can generally not be read off the diagrams.

5 Experiments

This section presents results for datasets generated with a software that simulates the spatial motion of cells. It serves as an illustration of our topological methods, and as a proof of concept. The results encourage us to proceed to the original goal of applying the method to cells of developing zebrafish embryos in the near future.

Cell segregation. Cell sorting involves both the segregation of a mixed population of cells with different properties into distinct domains, and the active maintenance of their segregated state. It has been described to occur in vivo in a wide variety of biological processes, such as the early embryonic development of multi-cellular organisms. The segre-

gation behavior can be studied in vitro – in artificial cell culture conditions – by mixing cells with different properties, and observing their subsequent sorting behavior. Such studies have been instrumental in revealing the cellular properties driving the segregation process. Different theories have been postulated and experimentally tested to explain the mechanism that drive cell sorting both in vitro and in vivo. However, to determine relative contributions of cellular properties requires the accurate recording as well as a quantitative and descriptive analysis of the sorting process in space and time. This is the purpose of this paper, namely a refined measurement of the spatial process that will allow us to compare different experimental conditions and quantify similarities and differences to wild type processes.

Here we use the zebrafish gastrulation as a case-study, in which mesoderm cells are induced and segregate from the overlying ectoderm germ layer. This process can be recapitulated in vitro by randomly mixing population of ectoderm and mesoderm cells in culture, and using advanced time-lapse microscopy imaging techniques [11] to observe how mesoderm cells engulf the ectoderm cells.

Simulations. Because of the difficulties of accurately tracking the trajectories of cell nuclei in real data, we concentrate on simulated data to demonstrate our methods. We use the publicly available simulation software *CompuCell3D*¹ [14] to get data sets that imitate the cell segregation process. The software uses a 3-dimensional version of the widely applied *Cellular Potts model* (also known as *Glazier and Graner model* [8]) to describe in vitro segregation of ectoderm and mesoderm cells. Each cell is represented by a set of voxels (unit integer cubes in \mathbb{R}^3), and randomly tries to extend into the empty space or a neighboring cell. An *elementary change* is the alteration of the membership status of a single voxel. The decision whether or not to accept such a change depends on an energy function that encodes the characteristics of the dynamic process. Evaluating the energy before and after the change, the difference is translated into the probability of acceptance. It is higher for negative differences, which drive the configuration toward smaller energy.

We give more details on our specialization and refer to [14] for the general simulation framework. The first term of our energy function constrains the shape of the cells. Setting the target volume to 49.0, each cell increases the energy by the square of its deviation from the target value. The same holds for the surface area, where we set the target to 26.0. The software allows for weights to control the relative influence of these terms, and we use weights 100.0 for the volume and 3.0 for the area. The second term of our energy function influences the inter-cellular behavior. We call two neighboring voxels an *interface* if they belong to different cells, or

to a cell and empty space. In our setup, an interface between two cells decreases the energy, so that cells tend to stick to each other. Specifically, we define two cell types, which we call *red* (simulating ectoderm) and *blue* (simulating mesoderm). An interface between two blue cells, or between a red and a blue cell decreases the energy by 200.0, while an interface between two red cells decreases the energy by 400.0. This results in red cells sticking to each other stronger than blue cells. Finally, an interface between a cell and empty space increases the energy by 20.0 for a blue cell, and by 100.0 for a red cell, with the effect that cells generally avoid contact with the surrounding space, and red cells do so with more determination than blue cells.

Datasets. We construct the initial configuration of our experiments by tiling the domain into $6 \times 6 \times 6 = 216$ cubes and placing one cell into each cube to fill it completely. This setup leads to cells that are larger than the target volume, resulting in a high initial value of the energy function. In the first steps of the simulation, the cells shrink and cluster in the middle of the available space. However, some of them lose contact and stay separated from the bulk for a long time, if not the entire time. We consider three simulations. In the

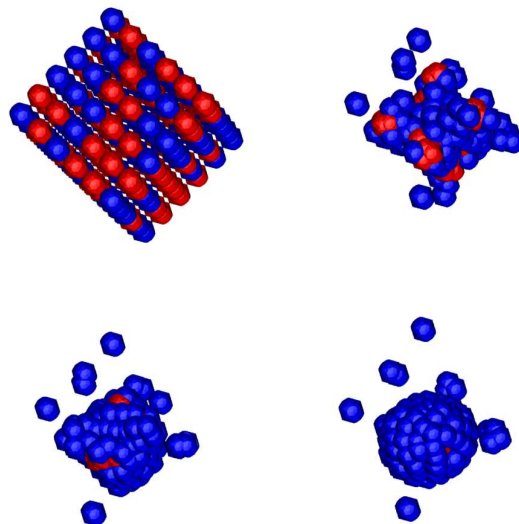


Figure 8: The restricted Voronoi tessellation at four moments in time. At the beginning, the cells form a cubical grid (upper-left). Next, the cells move toward the middle of the available space, leaving some outliers behind (upper-right). Thereafter, the blue cells begin to engulf the red cells (lower-left). Finally, the blue cells form a sphere surrounding a ball of red cells (lower-right).

first, we color each cell either red or blue according to a fair coin flip; see Figure 8. This intermixed configuration mimics the biologically interesting process of cell segregation. The second and third are control experiments in which all cells

¹<http://www.compuCell3d.org/>

are of the same type, either red or blue. To obtain trajectories, we compute the center of gravity of each cell at various moments in time and interpolate them piecewise linearly. In all experiments, we set the radius of the ball approximating one cell to $\alpha_0 = 4.0$, and we model the arrangement of cells by the thus defined restricted Voronoi tessellation.

As explained in the preceding sections, the behavior of the cell population in time is characterized by the persistence and image persistence diagrams of the time function restricted to the various medusas. There are too many diagrams to fit in this paper, so we show a subset, selected to illustrate interesting behavior of the data. For each diagram, we count the dots and we compute the 1-norm, which we define as

$$\|\text{Dgm}(f)\|_1 = \sum_{X \in \text{Dgm}(f)} \text{pers}(X). \quad (13)$$

These numbers are collected for each dimension and each subdiagram separately, giving a good overall impression of the activity recorded in the diagram.

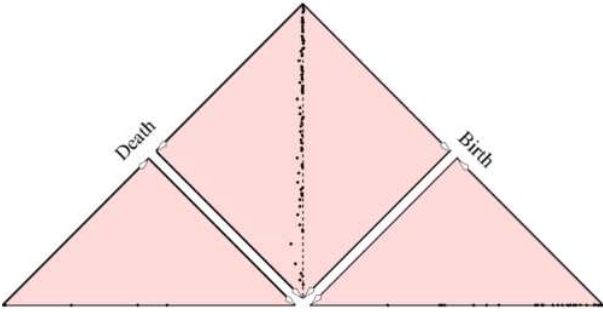


Figure 9: The 2-dimensional persistence diagrams of the restricted Voronoi medusa of **B**.

Multi-chromatic data. We start with the discussion of the intermixed simulation, letting **R** and **B** denote the corresponding sets of red and blue cells. See Figure 8 for illustrations of the restricted Voronoi tessellation, Figure 9 for one of the persistence diagrams, and Table 2 for the quantitative summaries of all diagrams, divided up into dimensions and subdiagrams. The only high-persistence feature in **R** is recorded in the horizontal subdiagram and simply represents the component itself. There are additional high-persistence features recorded in the 1-dimensional vertical and relative subdiagrams of **B**, representing the isolated (blue) cells that lose contact to the bulk during the initial phase. Cells recorded in the vertical subdiagram merge back into the bulk, while cells recorded in the relative subdiagram remain separate until the end of the simulation. In addition, the 2-dimensional horizontal subdiagram of **B** records features whose persistence adds up to a value of 0.76. They describe the sphere of blue cells that wrap around the red

cells. The fact that it is represented by several instead of a single dot indicates that the sphere is pinched from time to time during the simulation; compare with Figure 8.

To cast additional light on the process, we show results for the blue restricted Voronoi medusa included in the blue unrestricted Voronoi medusa; compare the two bottom panels of Table 2. We no longer see any persistent features in the vertical or relative subdiagrams because the outliers are merged with the bulk by the unrestricted medusa. Furthermore, the image persistence diagram features a single horizontal void with high persistence, which is the consolidation of the many low-persistence horizontal voids of the blue restricted Voronoi medusa.

R	Ord	Hor	Vcl	Rel	sum
gaps	17 0.03	1 1.00	7 0.03	0 0.00	25 1.07
tunnels	103 0.04	5 0.00	0 0.00	1 0.00	109 0.05
voids	21 0.00	0 0.00	2 0.00	0 0.00	23 0.00
sum	141 0.08	6 1.00	9 0.03	1 0.00	157 1.12
B					
gaps	10 0.02	1 1.00	22 1.58	3 2.55	36 5.15
tunnels	237 0.27	37 0.05	21 0.04	41 0.02	336 0.40
voids	32 0.00	61 0.76	0 0.00	0 0.00	93 0.76
sum	279 0.30	99 1.82	43 1.62	44 2.57	465 6.32
B, Vor					
gaps	0 0.00	1 1.00	0 0.00	0 0.00	1 1.00
tunnels	9 0.02	4 0.01	3 0.00	7 0.00	23 0.04
voids	1 0.00	1 0.95	0 0.00	0 0.00	2 0.95
sum	10 0.02	6 1.96	3 0.00	7 0.00	26 2.00

Table 2: From top to bottom: the sizes and 1-norms of the persistence diagrams of the red and blue restricted Voronoi medusas, and of the image persistence diagrams defined by the inclusion of the blue restricted Voronoi medusa in the blue Voronoi medusa. Bold-face numbers correspond to dots in Figure 9.

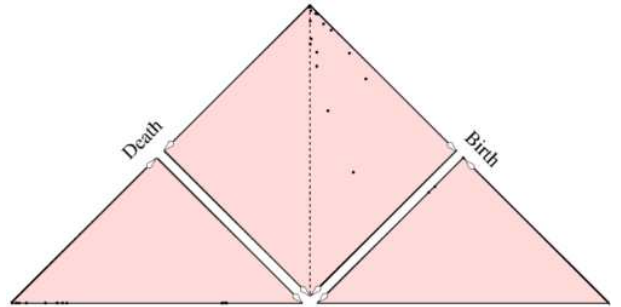


Figure 10: The 1-dimensional persistence diagrams of the restricted Voronoi medusa of **BB**.

Mono-chromatic data. The control experiments use simulations with cells of a single type, which is either red or blue. To avoid confusion with the data from the previous experiment, we denote the mono-chromatic sets of cells by **RR** and **BB**. Table 3 presents the quantitative summaries of the

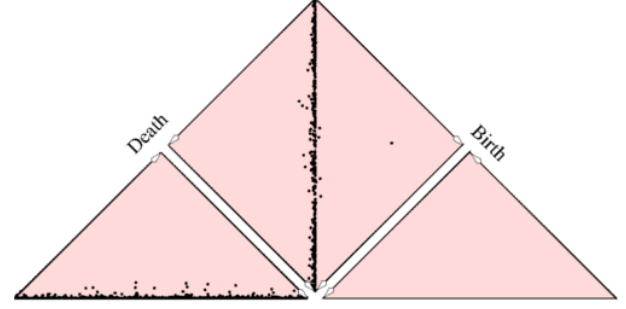
topological analysis. The process looks similar for both colors: there is a component formed by the majority of the cells (represented by a dot in horizontal subdiagram), and there are several outlier cells (represented by dots in the vertical and relative subdiagram). Moreover, many low-persistence ordinary tunnels are created in the initial phase, whose persistence adds up to a significant number.

There are also differences. First, the blue cells create many vertical gaps (components that split and re-merge during the simulation), while they are rare for the red cells. An inspection of Figure 10 reveals that most of these splits happen early in the simulation, when the cells still shrink. This effect can be explained by the smaller stickiness of the blue cells, which makes it more likely for them to lose contact with the bulk. Once separated, outliers move randomly in space and find contact with the bulk only by chance. Curiously, the red medusa has the same number of relative gaps (components that split and the pieces stay separate) as the blue medusa. It suggests that the red outliers find it more difficult to merge back into the bulk. One reason may be their stickiness to each other, and their phobia of the empty space, which limits their search activity.

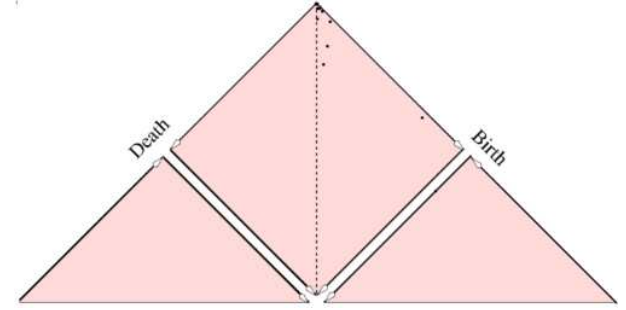
A second difference can be observed in the vertical voids, which only exist for the red cells. Recall that they are created by puncturing and destroyed by filling. As we can read off the table, these holes are short-lived. There is no biological reason why cells would create such holes, and they should indeed be considered an artifact of the abstraction of cells as relatively inflexible restricted Voronoi regions. Nevertheless, even this artifact tells us something about the simulation process. The presence of such holes means that the red cells tend to deviate more from the shape of a round ball than the blue cells. This is consistent with the energy function in our setup in which red-red interfaces are more favorable than blue-blue interfaces. Therefore, the red cells get more reward for increasing their interface area at the expense of deviating from the target volume and surface area.

RR	Ord		Hor		Vcl		Rel		sum	
gaps	0	0.00	1	1.00	2	0.01	3	2.99	6	4.01
tunnels	607	0.60	47	0.04	1	0.00	3	0.00	658	0.64
voids	339	0.03	0	0.00	24	0.02	10	0.00	373	0.06
sum	946	0.64	48	1.04	27	0.04	16	2.99	1037	4.72
BB										
gaps	0	0.00	1	1.00	15	1.52	3	2.56	19	5.08
tunnels	582	0.55	81	0.09	0	0.00	16	0.00	679	0.65
voids	310	0.02	0	0.00	0	0.00	4	0.00	314	0.02
sum	892	0.58	82	1.09	15	1.52	23	2.56	1012	5.76

Table 3: The sizes and 1-norms of the persistence diagrams of the red medusa at the top and the blue medusa at the bottom. Boldface numbers correspond to dots in Figure 10.



(a) Image diagram induced by the inclusion of the restricted Voronoi medusa of B_1 in unrestricted Voronoi medusa of the same set of cells.



(b) Image diagram induced by the inclusion of the restricted Voronoi medusa of B_1 in the restricted Voronoi medusa of BB .

Figure 11: The 1-dimensional image persistence diagrams of the inclusion of B_1 in its corresponding Voronoi medusa at the top, and the blue combined medusa at the bottom.

Image diagrams. We create a randomly intermixed process by decomposing BB into two sets of cells, B_1 and B_2 , with fair coin flips. There is no reason to expect a sorting process similar to the simulation with red and blue cells. Instead, the two (identical) cell types form a random pattern with arbitrary neighbor exchanges over time.

The quantitative results for B_1 are given in Table 4; those for B_2 are similar. Note the large number of tunnels, which are distributed over the entire simulation. Recall that we have two possibilities for computing image persistence: by mapping the restricted Voronoi medusa of B_1 into the unrestricted Voronoi medusa of the same set of cells, or into the restricted Voronoi medusa of BB . For comparison, we display the quantitative results of both inclusions in Table 4, showing the corresponding 1-dimensional diagrams in Figure 11. The majority of features – both in terms of number and persistence – carry over to the unrestricted Voronoi medusa, which suggests that most of the holes in the medusa of B_1 are caused by the presence of cells in B_2 . This is consistent with the relative sparsity of the image diagrams defined by the inclusion of B_1 in BB . Notable exceptions are the gaps recorded in the 1-dimensional vertical and relative subdiagrams, which describe the outliers separated from the bulk mentioned earlier.

B ₁	Ord		Hor		Vcl		Rel		sum	
gaps	11	0.00	1	1.00	42	1.28	1	0.76	55	3.05
tunnels	1406	4.32	629	5.41	648	6.02	1273	4.60	3956	20.36
voids	32	0.00	0	0.00	0	0.00	0	0.00	32	0.00
sum	1449	4.33	630	6.41	690	7.31	1274	5.36	4043	23.42
<hr/>										
B ₁ , Vor										
gaps	0	0.00	1	1.00	10	0.66	0	0.00	11	1.66
tunnels	1113	4.05	445	3.70	627	5.92	1183	4.47	3368	18.15
voids	0	0.00	2	0.00	0	0.00	0	0.00	2	0.00
sum	1113	4.05	448	4.70	637	6.58	1183	4.47	3381	19.82
<hr/>										
B ₁ , BB										
gaps	0	0.00	1	1.00	15	1.04	1	0.76	17	2.80
tunnels	149	0.11	15	0.01	0	0.00	1	0.00	165	0.13
voids	32	0.00	0	0.00	0	0.00	0	0.00	32	0.00
sum	181	0.11	16	1.01	15	1.04	2	0.76	214	2.94

Table 4: From top to bottom: the sizes and 1-norms of the persistence and image persistence diagrams of the first blue medusa. Boldface numbers correspond to dots in Figure 11.

6 Discussion

This paper introduces the persistent homology analysis of a medusa as a novel method to measure cell segregation. To provide a proof of concept, we have computed these measurements for simulated time series of 3D data. The application of the method to imaged cell segregation in zebrafish embryos is forthcoming.

The medusa introduced in this paper is related to the vineyard described in [6, Section VIII.1]; see also [3]. However, there are differences. Specifically, the vineyard would be constructed for two parameters, the restricting radius, α , and the time, t . The result is a richer structure, namely a collection of curves in $\mathbb{R}^2 \times [0, 1]$, which is therefore more difficult to understand. In contrast, the medusas fix α to α_0 and in this way facilitate a more compact representation of a subset of the vineyard. This is appropriate for biological cells whose size does not vary substantially with time, and it gives topological information that is easier to interpret and more directly relevant to the object of study. However, we need to keep in mind that with this choice, the persistence diagrams are not stable under the bottleneck distance. The results are particularly sensitive to the radius, which leaves holes if chosen too small and absorbs holes if chosen too big.

The work in this paper has motivated the extension of Alexander duality from spaces to functions, as proved in [7]. In particular, the Euclidean Shore Theorem in that paper states that the persistence diagram of the time function restricted to the boundary of a Voronoi medusa can be obtained from the diagram of the time function on the Voronoi medusa. Generically, the boundary is a 3-manifold without boundary, so that the time function can be understood in Morse theoretic terms, which is sometimes convenient.

In conclusion, we note that the framework introduced in this paper applies to general point processes that unfold in time. The latter model a variety of problems of which

we mention a few: molecules of two fluids mixing after a shock-wave; microbes forming microfilms; a flock of birds getting into formation; two teams competing in a soccer game; galaxies moving under the influence of gravity. The measurements taken within this framework are significantly coarser than what could be said by studying braids [12] or the loop spaces of configuration spaces needed to detect topological differences for particles moving in 3 dimensions [1]. This is not necessarily a disadvantage since the coarser information is easier to compute as well as to comprehend.

References

- [1] F. R. COHEN AND S. GITLER. On loop spaces of configuration spaces. *Trans. Amer. Math. Soc.* **354** (2002), 1705–1748.
- [2] D. COHEN-STEINER, H. EDELSBRUNNER, J. HARER AND D. MOROZOV. Persistent homology for kernels, images, and cokernels. In “Proc. 20th Ann. ACM-SIAM Sympos. Discrete Alg., 2009”, 1011–1020.
- [3] D. COHEN-STEINER, H. EDELSBRUNNER AND D. MOROZOV. Vines and vineyards by updating persistence in linear time. In “Proc. 22nd Ann. Sympos. Comput. Geom., 2006”, 119–126.
- [4] H. EDELSBRUNNER. Deformable smooth surface design. *Discrete Comput. Geom.* **21** (1999), 87–115.
- [5] H. EDELSBRUNNER. *Geometry and Topology for Mesh Generation*. Cambridge Univ. Press, Cambridge, England, 2001.
- [6] H. EDELSBRUNNER AND J. L. HARER. *Computational Topology. An Introduction*. Amer. Math. Soc., Providence, Rhode Island, 2010.
- [7] H. EDELSBRUNNER AND M. KERBER. Alexander duality for functions: the persistent behavior of land and water and shore. In “Proc. 28th Ann. Sympos. Comput. Geom., 2012”, 249–258.
- [8] F. GRANER AND J. A. GLAZIER. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.* **69** (1992), 2013–2016.
- [9] C.-P. HEISENBERG AND G. KRENS. Cell sorting in development. *Current Topics in Developmental Biology*, to appear.
- [10] M. KERBER AND H. EDELSBRUNNER. The Medusa of spatial sorting: 3D kinetic alpha shapes and implementation. Manuscript, IST Austria, Klosterneuburg, Austria, 2012.
- [11] A. V. KLOPPER, S. G. KRENS, S. W. GRILL AND C.-P. HEISENBERG. Finite-size corrections to scaling behavior in sorted cell aggregates. *Europ. Phys. J. E, Soft Matter* **33** (2010), 99–103.
- [12] W. MAGNUS. Braid groups: a survey. In “Proc. 2nd Intl. Conf. Theory Groups”, Springer Lect. Notes Math. **372**, 463–487, 1974.
- [13] J. R. MUNKRES. *Elements of Algebraic Topology*. Perseus, Cambridge, Massachusetts, 1984.
- [14] M. H. SWAT, S. D. HESTER, R. W. HEILAND, B. L. ZAITLEN, J. A. GLAZIER AND A. SHIRINIFARD. *CompuCell3D Manual and Tutorial Version 3.6.1*. Biocompl. Inst. and Depart. Physics, Indiana University, Bloomington, Indiana, 2012.